Introduction to Viruses

Introduction

Viruses are infinitely small compared to a cell. They are the simplest form of life that exists, and viruses' status as "alive" is debatable. Viruses generate no energy. The nucleotides and proteins of viruses require the presence of cellular machinery to generate their energy. A virus does nothing until it encounters a cell it recognizes. The virus gets in the cell, and still does nothing. The viral genome is nucleotides. The host cell machinery recognizes nucleotides. Ribosomes read mRNA in the cytoplasm. Polymerases read and synthesize nucleotides in the nucleus. The virus lets the host machinery do to its nucleotides what the host cell normally does to all nucleotides. But it just so happens that what the host machinery does is make more virus. Sometimes viruses also bring along enzymes of their own. But all of virology comes down to do replication, transcription, and translation.

Which means that viruses can be incredibly simple to master if first you master biochemistry. This Virus series is part of the Microbiology module. It expects you to have completed *Biochemistry: DNA to Protein* and to be familiar with the processes of replication, transcription, and translation. We give you a little refresher to get started in case you have not completed that, or are a little rusty. The goal of this lesson is to introduce the vocabulary of viruses, speak generally about the organization of the lessons, and orient you to the rest of the module. The next lesson will tackle the life cycle of viruses. With the physiology and biochemistry out of the way in the first two lessons, we can then focus on the diseases the viruses cause in lessons 3 through 7. We conclude the series with a lesson on antivirals.

Biochemistry Review

Humans have double-stranded DNA. DNA stands for deoxyribonucleic acid, the OH group at the second position on the pentose sugar having been removed. DNA has 4 nucleotides—A pairs with T, G pairs with C. Humans can **replicate** their DNA using each strand of DNA as a template. In **replication**, **DNA polymerase** reads the template DNA and assembles a chain of base pairs, complementary and antiparallel to the template strand. Because both template strands are complementary and antiparallel to each other, and the replicated strands are complementary and antiparallel to each other, the result is two identical double strands of DNA.

Humans have single-stranded RNA, called messenger RNA. RNA stands for ribonucleic acid, an OH group at the second position on the pentose sugar. RNA has 4 nucleotides—A pairs with U, G pairs with C. Humans **transcribe** their DNA using one strand, the template strand, to build an RNA strand that is complementary and antiparallel to the template strand. Because the template strand (the one being used to make the RNA) is complementary and antiparallel to the coding strand, the mRNA that is made is exactly the same as the coding strand, except there are U's in RNA where there are T's in the DNA. **Transcription** is carried out by **RNA polymerase**. The mRNA is dispatched from the nucleus to the cytoplasm, where ribosomes are waiting. Ribosomes read the mRNA and **translate** the mRNA into an **amino acid sequence**; protein is made from mRNA.

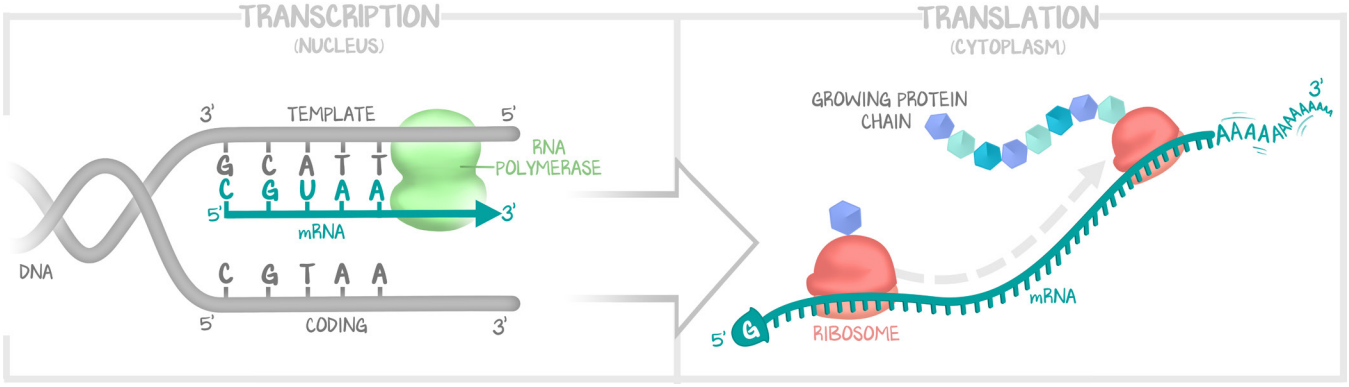

Figure 1.1: Biochemistry Review
Transcription uses RNA polymerase to construct an RNA transcript that will be processed into messenger RNA. RNA polymerase reads the template strand, constructing the mRNA from its 5' end towards its 3' end. The transcript, the mRNA, is complementary and antiparallel to the template strand. The template strand is complementary and antiparallel to the coding strand. Therefore, the transcript and the mRNA carry same the message, except that RNA uses uracil (U) and DNA uses thymine (T). Always orient your code-that-matters 5' left, 3' right. Transcription happens in the nucleus. Translation occurs in the cytoplasm. The mature mRNA has a methylguanine gap on the 5' end and a poly-A tail on the 3' end. Ribosomes bind to start codons on the 5' end and move towards the 3' end, adding amino acids, one codon, one 3-nucleotide long string, at a time.

Orientation matters—for the enzymes and for you. In biochemistry, we said always to put the strand that matters 5' left, 3' right, meaning the start is on the left, and the end is on the right, the way you are reading right now. The coding strand is 5' left, 3' right, the mRNA is 5' left, 3' right, and so the amino acid sequence is N-terminus left, C-terminus right. Ribosomes can read the mRNA one way only—starting at the 5' end and reading toward the 3' end. RNA polymerase can build one way only, starting at the 5' end of the new strand, building toward the new strand's 3' end. DNA polymerase can build one way only, starting at the 5' end of the new strand, building toward the new strand's 3' end.

Viral Enzyme Terminology

Because viruses are tricksy and hijack host cell machinery, we have to be more specific with how we name enzymes. In discussing normal human function, "DNA polymerase" and "RNA polymerase" assume that DNA is being read so the name of the enzyme describes what it builds. That assumption works in discussing normal human function because that's the only direction human cells go—DNA is read and a complementary strand is made. But studying viruses necessitates being clearer. We use names to specify both what is read ("-dependent") and what is made ("-polymerase").

Human DNA polymerase reads DNA and makes complementary DNA. Therefore, under our viral nomenclature, it is named **DNA-dependent** (reads DNA) **DNA polymerase** (makes DNA). DNA-dependent DNA polymerase reads a strand of DNA and makes a strand of DNA. Similarly, human RNA polymerase reads DNA and makes a strand that is complementary and antiparallel made of RNA—the mRNA transcript. Therefore, RNA polymerase is named **DNA-dependent** (reads DNA) **RNA polymerase** (makes RNA). If a virus is made of DNA, all it needs to do is get to the host nucleus. Being made of DNA, the virus can just let the host machinery do what it does—DNA-dependent DNA-polymerase (what you know as DNA polymerase from biochemistry) replicates the virus. DNA-dependent RNA polymerase (what you know as RNA polymerase from biochemistry) transcribes the DNA virus into mRNA, which is sent to the cytoplasm to be transcribed into viral proteins.

DNA viruses use host machinery to replicate and transcribe viral proteins. Host machinery, normal human enzymes, read DNA and make either DNA or RNA.

If a virus is RNA it needs only get to the cytoplasm. Because the virus is RNA, ribosomes in the cytoplasm start translating the virus into viral proteins. But the virus is RNA. Human cells don't have enzymes that read RNA. So in order to make more virus, the virus will either need to bring with it into the cell, or code for, a protein for ribosomes to make, that **reads RNA** and **makes RNA**. Reads RNA (RNA-dependent) makes RNA (RNA polymerase). Viral **RNA-dependent RNA polymerase** is how most RNA viruses replicate. Retroviruses (such as HIV) are RNA viruses that become DNA viruses in host cells. To do that, they must read RNA and make DNA. RNA-dependent DNA polymerase is **reverse transcriptase**.

Sounds awfully burdensome to have to write out DNA-dependent DNA polymerase every time. So we shorten it with convention. As shown in Figure 1.2, we use the model Nucleotide-d-Nucleotide-p, with d for dependent and p for polymerase. If it is DNA, the letter for the nucleotide is D; for RNA the letter for the nucleotide is R.

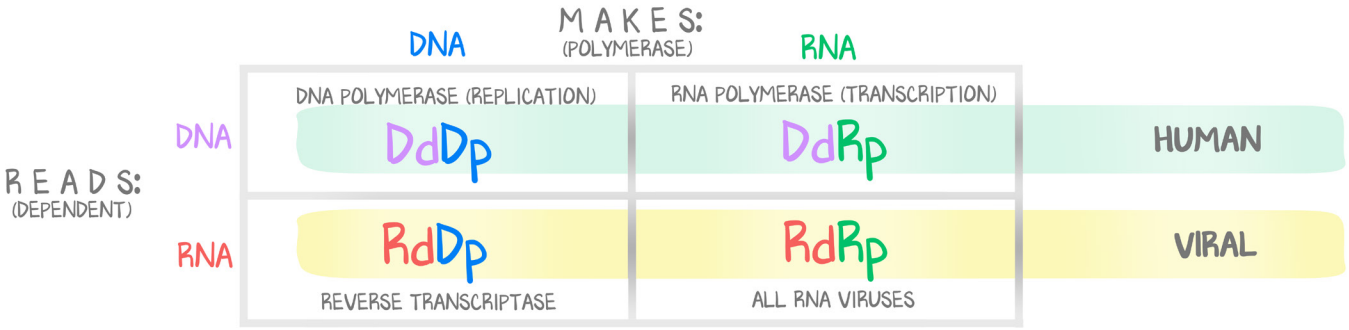

Figure 1.2: Enzyme Terminology
In human biochemistry we used only DNA polymerase and RNA polymerase because normal human polymerase activity only reads DNA and either makes DNA (replication) or RNA (transcription). Viruses also need to replicate and have human ribosomes make proteins for them. Viruses can be either DNA or RNA. Therefore, we need an easy way to describe the action taking place. A polymerase will read either DNA or RNA, and is assigned "dependent" on what it reads. A polymerase can make DNA or RNA, and is assigned "polymerase" for what it makes. DdDp = DNA-dependent DNA polymerase, DdRp = DNA-dependent RNA polymerase, RdDp = RNA-dependent DNA polymerase, RdRp = RNA-dependent RNA polymerase.

Throughout this course we will sometimes use the abbreviation and sometimes use the enzyme spelled out. When it is written out, say it out loud, slowly.

This took a long time to get through. And if you've followed along, there are only four permutations. So why all the fuss? Remember Wrong Way #1 from Bacteria #5: *OME Taxonomy*? Go check it out if you don't. Microbiology is about **clustering**. Rather than recite for every DNA virus, "uses host DdDp and DdRp, does not need to bring its own," we can say "All DNA viruses use host DdDp and DdRp." Only RNA viruses are going to use RdRp or RdDp.

Virus, Virion, and Terminology

Technically the word "virus" means what we are going to call genome, a string of nucleotides. What we commonly think of as a virus is technically a virion, the combination of the genome and structural proteins. But colloquially, inside and outside of medicine, we throw around the word "virus" to mean the infectious element, the thing causing the symptoms. It's also more natural, easier to read and say. From this point forward, virus means virion. Genome means nucleotides. That way we can mean the concept virus, the persona virus, or the structure virus—using the word the way we have been trained our entire lives. A **virus** is the thing that can go infect another cell, the **genome** is the string of nucleotides, and the **capsid** is the structural proteins. The combination of a capsid (the container) and the genome (the nucleotides) is called a **nucleocapsid**.

All viruses consist of a genome and a capsid. All the genome can do is be read. It doesn't survive outside of the virus or a host cell, it doesn't have a shape, cannot interact with cells. The genome is the lifeless instructions on how to make more virus. The **genome** also encodes for capsid proteins and any other viral proteins this particular virus might need. The **capsid** is made of structural proteins made by the host organism and is encoded by the genome. The capsid gives the virus its shape. Viruses can assume one of two shapes—icosahedral or helical. The **icosahedral** shape contains the genome within a little forcefield. This forcefield allows the virus to exist outside of a host cell. The **helical** shape offers no such protection, and is unstable outside of a host cell.

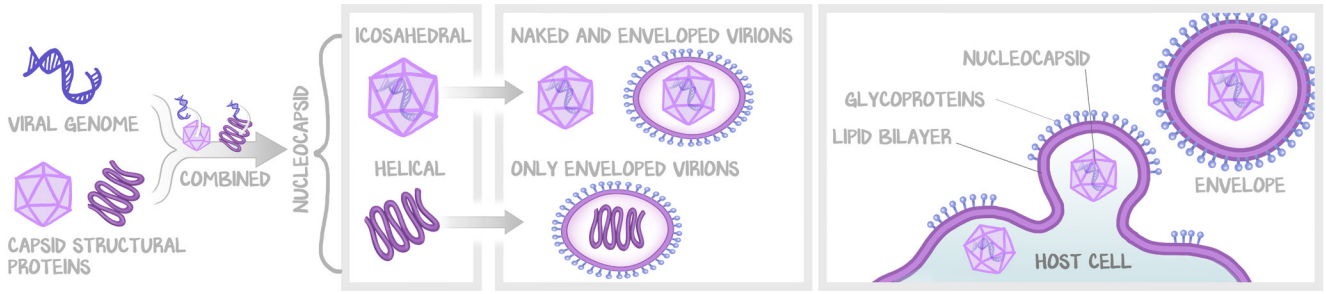

Figure 1.3: Viruses
The combination of the viral genome and its capsid is the nucleocapsid. Some nucleocapsids are icosahedral in shape, some are helical. All helical nucleocapsids get wrapped in an envelope. Some icosahedral nucleocapsids get wrapped in an envelope, some remain naked.

Some viruses have **envelopes**. An envelope is the plasma membrane of the host cell, modified with viral glycoproteins to serve the virus. When an enveloped virus leaves a cell, it leaves wrapped in that host-plasma-membrane-with-viral-proteins-in-it, budding off from the host plasma membrane. Only those viruses which code for proteins that embed themselves within the plasma membrane (synthesized in the RER, processed by Golgi, etc.) can be enveloped. These **viral membrane proteins** not only identify that region of host plasma membrane as viral, but they also usually aid in infecting the next cell. Because the envelope is plasma membrane and the virus starts inside the cytoplasm as it gets enveloped, there is also cytoplasm around the virus within the envelope, space between the nucleocapsid and the membrane itself. That means enveloped viruses have some space to carry with them any enzymes they might need. **Viral cytoplasmic proteins** are required for that virus to start replicating in the next host cell.

A virus without an envelope is called a **naked virus**. Only icosahedral-shaped viruses can be naked. All helical viruses must be enveloped. A virus with an envelope is called an **enveloped virus.** The envelope has viral proteins (aka foreign proteins) embedded in a plasma membrane. Our immune system is primed for that arrangement. By providing foreign proteins, the virus does enhance its ability to infect the next cell, but it also gives the immune system a target for antibodies. Naked viruses have fewer identifiable features, so naked viruses tend to evade the immune system better. The envelope stores viral proteins in the cytoplasm. This is almost always because the virus needs those viral proteins to do anything in the next cell it infects. What that also means is that if the envelope were compromised, the enzymes the virus needs would be lost. Cell membranes are also lipid, so dissolve in detergents. Enveloped viruses may resist osmotic forces better than a naked virus, but enveloped viruses are usually easier to clean from surfaces. There is a tradeoff—get cool toys (enveloped virus, with their proteins), or go undetected (naked).

More about the Genome

The virus genome can be either DNA or RNA, but never both.

DNA viruses get into a host cell and go to where DNA is processed in the host cell, the nucleus. There, the host has the equipment the DNA virus needs to make more of itself (DdDp, replication) and to make the proteins it will need to move on to the next cell (DdRp, transcription). DNA viruses are always double stranded (except for parvovirus). Therefore, **DNA viruses** are **double stranded**, replicate in the **nucleus** using **host DdDp**, and transcribe mRNA to make the proteins they'll need using **host DdRp**. They don't need their own machinery to come with them. That means they don't need an envelope, so can be a naked virus.

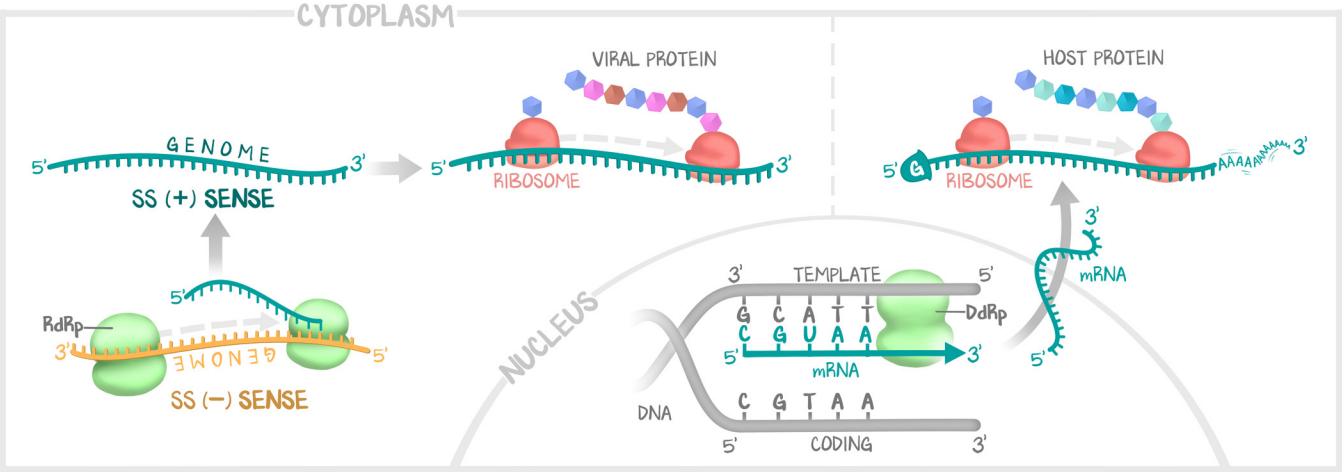

Figure 1.4: Sense and Antisense
Ribosomes know to start on the 5' end of mRNA and read toward the 3' end. It's the only thing they know how to do. If a viral genome "makes sense" to a ribosome—the code of the genome is 5' left to 3' right—it is said to be a positive-sense genome. If a viral genome doesn't "make sense" to a ribosome—the code of the genome is 3' left to 5' right—it is said to be negative sense or antisense. DNA viruses, which are double-stranded and operate from the nucleus, need no such distinction.

RNA viruses get into a host cell and go where RNA is processed in the host cell, the cytoplasm. Normal transcription produces an mRNA that is oriented 5' left, 3' right. Ribosomes recognize this orientation and start translation on the 5' end, moving to the 3' end. The correct protein is made. mRNA makes sense to a ribosome, and the protein they translate is the one coded by the original DNA. RNA viruses are always **single stranded** (except reovirus). The viral genome codes for the proteins it will need, and needs host ribosomes to translate that code. Viral genomes that are ready to be translated must make sense to the ribosome. Viral genomes that are oriented 5' left to 3' right are called **positive-sense viruses**, abbreviated ss(+)RNA for single-stranded positive-sense RNA. These viruses do not need any additional equipment, just like DNA viruses. Their code will be translated into the proteins needed for replication, **RdRp**. Viral genomes that are oriented 3' left 5' right do not make sense to ribosomes and are called **negative-sense viruses**, abbreviated ss(−)RNA for single-stranded negative-sense RNA. These viruses must first use themselves as the template strand to make a strand of RNA the ribosome understands. These viruses must bring RdRp with them. RNA viruses are therefore **single stranded**, replicate in the **cytoplasm,** and are the code for translation. ss(+)RNA viruses use only host machinery; ss(−)RNA viruses must bring with them a viral protein.

Taxonomy and Overview

Here, we lay out the map for the rest of the course and take advantage of the concepts from above.

DNA viruses are always icosahedral. DNA viruses CAN use only host machinery; therefore DNA viruses can be naked. Not all naked viruses are DNA, and not all DNA viruses are naked. This implies only that they can be naked.

ss(+)RNA viruses are always icosahedral (except coronavirus). ss(+)RNA viruses CAN use only host machinery; therefore ss(+)RNA viruses can be naked. Not all naked viruses are RNA, and not all ss(+) RNA viruses are naked. This implies only that they can be naked.

All ss(–)RNA viruses are helical. All helical viruses must be enveloped. All ss(–)RNA viruses must bring with them RdRp to initiate replication in the cytoplasm. All ss(–)RNA viruses are enveloped.

All DNA viruses are double stranded (except parvovirus). All RNA viruses are single stranded (except reovirus, which we will not study).

All DNA viruses replicate in the nucleus. All RNA viruses replicate in the cytoplasm.

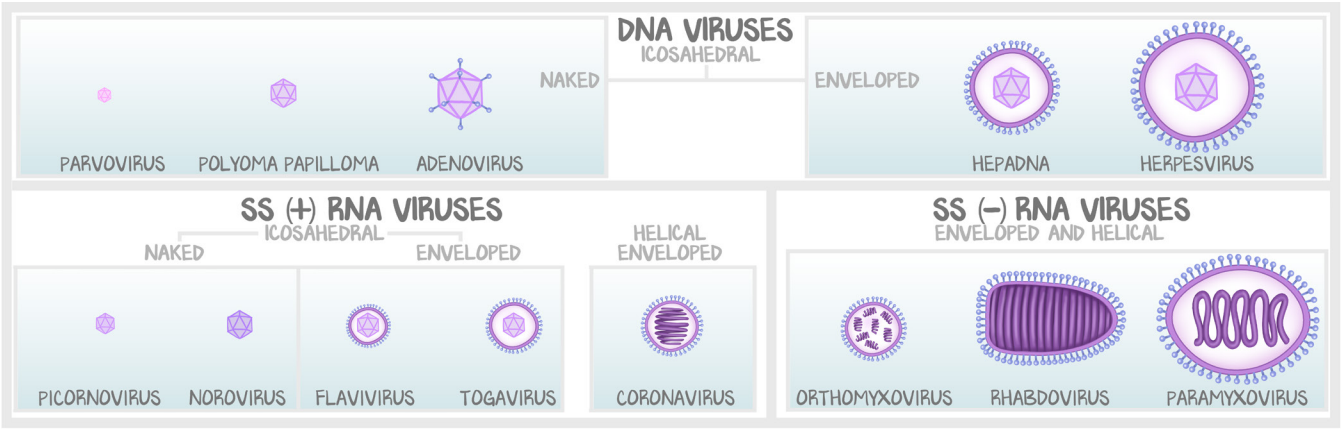

Figure 1.5: Viral Taxonomy and the Plan of the Course
This table is the organizer for the rest of the Viruses series. We will discuss DNA viruses, ss(+)RNA viruses, and ss(–)RNA viruses in subsequent lessons. This table is also stripped of many viruses commonly seen in virology textbooks. We are not teaching you every virus. We are teaching you the viruses you need to know. Use this table as a reference. Take notes on it, or redraw it yourself on a separate piece of paper, adding to it as you progress through the series.

Because our course is laid out DNA viruses, ss(+)RNA viruses, and then ss(–)RNA viruses, and you will be learning the viruses amongst other viruses like themselves, contained within discrete lessons, it becomes easier to memorize. The advance organizer is the lesson—since you learn the viruses with similar microbiology together, we also lighten that memorization barrier by showing you what you actually have to remember and what you can infer. The lesson contents are revealed by the lesson name: #3: *DNA Viruses*, #4: *ss(+)RNA Viruses*, and #5: *ss(–)RNA Viruses*. Where we break that mold is in the lessons that follow.

	DNA	SS(+)RNA	SS(-)RNA
Nucleotide	dsDNA (except parvovirus)	ssRNA	ssRNA
Nucleocapsid shape	Icosahedral	Icosahedral (except coronavirus)	Helical
Replication	Nucleus	Cytoplasm	Cytoplasm
Envelope	Herpesvirus and hepadnavirus have an envelope	Flavi-, toga-, corona- are enveloped	All must be enveloped
Host enzyme	DNA-dependent RNA polymerase, on entry	Ribosomes, on entry	Ribosomes, after viral RNA-dependent RNA polymerase

Table 1.1: Simplifying Memorization
This table is read: "all DNA viruses are double stranded, except parvovirus. All DNA viruses are icosahedral. All DNA viruses replicate in the nucleus. Some DNA viruses are enveloped; they are the herpesvirus and hepadnavirus. The rest are naked. All DNA viruses use host DdRp." Now, you try the next two columns. Unless you're in a quiet place around other people, do it out loud. But actively mouth the words. You'll feel silly. You'll also start viruses on the right foot.

What about HIV?

HIV is special and breaks all the rules. We cover HIV in detail together with the antivirals to treat it in its own lesson, Viruses #6: *HIV & AIDS*. It doesn't fit into our nicely packaged and simplified taxonomy, so we pull it out.

HIV is an RNA virus that replicates in the nucleus because it uses reverse transcriptase to build DNA from its genome. The DNA synthesized from viral RNA is then used by the host just like regular DNA. Reverse transcriptase is RNA-dependent DNA polymerase.

What about Hepatitis?

Hepatitis is inflammation of the liver. There are five viruses that cause hepatitis. They occupy a spattering of DNA and RNA classifications, so are taught in the context of the disease hepatitis (Viruses #7: *Hepatitis Viruses*), where we compare their genomes and specific mechanisms of actions. As you will see, even though each hepatitis virus shares features of its true viral family, removing them from the traditional taxonomy streamlines the mastery of the remaining viruses in each category. And teaching the hepatitis viruses together streamlines the understanding of those viruses.

270+: Antiviral Host Response, aka Why You Feel So Bad and Get Symptoms

The virus itself doesn't do too much to you. It just gets in your cells, uses them up, and jumps ship. It wouldn't cause very many symptoms, at least, until the organ it was using up died. Instead, **why you get symptoms** is because of the **host immune response**.

Interferon proteins (IFN-α and IFN-β) are secreted by infected cells to induce an antiviral response by neighboring cells. They are also the **cry for help** sent to **natural killer cells**. We didn't talk much about NK cells in Immunology, other than to say, "they survey for cancer and viruses." Well here they are, surveying for viruses. They know a cell is infected with a virus when the infected cell tells them it's infected—using interferons. It is **interferons** that give you **flu-like symptoms**. Natural killer cells kill infected cells. Interferons are released from a virally infected cell, which in turn tells nearby cells to watch out for viruses. Interferon has some mechanisms you don't need to know that result in an increase

in RNA endonuclease (kills RNA virus, chews up viral-coded mRNA) and decrease in elongation factor-2 (limiting translation). Interferon limits translation of viral protein, but it also limits translation of normal protein, too. And that's why you feel like crap when you get sick.

Recall also from Immunology that there are a number of host immune system responses that can facilitate fighting against viruses. **Cell-mediated immunity** is a combination of phagocytes (macrophages) and CD8 cytotoxic T cells. When antigen-presenting cells use MHC-2 to present antigens from intracellular pathogens destroyed by NK cells, a Th0 naïve CD4 cell becomes Th1, which in turn secretes IFN-γ (macrophages) and IL-2 (CD8 T cells). Independently, CD8 cytotoxic cells survey host MHC-1 self-antigens and release death cytokines to kill infected host cells who don't say the safe-word. Cell-mediated immunity is particularly good for killing intracellular infections. That isn't always good. The hemorrhagic fever of yellow fever and the extreme pain from dengue fever are caused by the death of organs. But it isn't the virus doing that, it's the host immune response to the virus.

Antibody-mediated immunity controls viremia (virus in the bloodstream). Regardless of the type of virus, **antibodies neutralize viruses**. The antibody can recognize viral glycoproteins on the envelope or on a naked capsid. As we learn in the next lesson, viruses must first attach to the cell membrane of a target cell. They do that by using **viral glycoproteins** to interact with **surface proteins on host cells**. If there is an antibody in the way, that attachment fails. Antibody-mediated immunity is particularly good for extracellular viruses. Not all antibody-mediated immunity is good, however. Cryoglobulinemia, an immune deposition disease (type III hypersensitivity), can occur in Hep B infection.